Ecological Modelling 377 (2018) 26–34

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Evaluation of CLM-Crop for maize growth simulation over Northeast China

Meiling Sheng[a,b,c], Junzhi Liu[a,b,c,*], A-Xing Zhu[a,b,c,d], David G. Rossiter[a,e], Liming Zhu[a,b,c], Guoqiang Peng[a,b,c]

[a] Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing, Jiangsu 210023, China
[b] State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing, Jiangsu 210023, China
[c] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, Jiangsu 210023, China
[d] Department of Geography, University of Wisconsin-Madison, Madison, USA
[e] School of Integrative Plant Sciences, Section of Soil & Crop Sciences, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

Keywords:
Community land model
Maize
Phenology
Crop yield
Northeast China

ABSTRACT

Climate change has significant impacts on crop yields and could greatly affect global food security. As an advanced process-based land surface model, the Community Land Model (CLM) includes a comprehensive crop model (CLM-Crop), which aims to simulate crop growth globally. Although this model has been evaluated in the United States, it is not clear whether it can be applied successfully in other parts of the world. In this study we evaluate the applicability of CLM-Crop for maize growth simulation in Northeast China, one of the major agricultural production areas in China. Simulated LAI, plant carbon, phenology and yields of maize were compared with observations at agricultural experiment and meteorological stations and with statistical reports. The CLM-Crop model overestimated LAI and underestimated leaf and stem carbon during the growing period. Planting and harvesting dates were overestimated in the eastern part of the study area, but underestimated in the southern part. Correlation ($r = 0.26$) between simulated and reported yield was poor. Yields were generally overestimated especially in the east and south parts, which may be due to imperfect farming practices on working farms. Some parameters, including temperature thresholds of planting and crop management parameters, have spatial heterogeneity rather than the default fixed parameters in CLM-Crop. Development of gridded crop parameters is expected to improve simulation of crop phenology and yield estimation at the regional and global scales.

1. Introduction

Climate change may significantly affect crop yields and therefore global food security (Peng et al., 2004; Rosenzweig et al., 2014; Schlenker and Roberts, 2009). Numerous studies have shown that climate change, such as increasing temperature and frequency and severity of extreme climate events, may have substantial effects on crop growth and yield. Actions must therefore be taken to adapt agriculture to climate change (Deryng et al., 2014; Lobell et al., 2014, 2012; Osborne et al., 2009; Siebert et al., 2014). In order to make correct decisions for agricultural adaptation, crop yield prediction under climate change scenarios is key information (Intergovernmental Panel on Climate Change (IPCC, 2007; Piao et al., 2010; Wang et al., 2011; Yao et al., 2015).

Crop yields under climate change are commonly predicted by two approaches: (1) empirical-statistical models (Bondeau et al., 2007; Fader et al., 2010) and (2) process-based crop models (Liu et al., 2007; Yao et al., 2015). Empirical-statistical models use regression or machine learning methods to establish the relationships between crop yield and climate variables based on historical observations (Ed et al., 2013; Hoffman et al., 2018; Lobell and Ortizmonasterio, 2007); these models are then extrapolated to plausible scenarios. These models are relatively straightforward and easy to construct, but they usually use temporally-aggregated climate data. They may thus miss important heterogeneities and may fail to capture the details of interactions between climate change and crop growth (He et al., 2015; Hoffman et al., 2018; Lobell and Asseng, 2017). Process-based crop models dynamically simulate the key biophysical and biochemical processes of crop growth, usually on a daily time step. So they better reflect the details of the climate and can represent the complex relationships between crop growth and climate change (Estes et al., 2013; Lobell and Asseng, 2017; Roudier et al., 2011; Zhao et al., 2013).

In the early stage of their development, process-based crop models were applied at the field scale where environment factors can be

* Corresponding author at: Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing, Jiangsu 210023, China.
E-mail address: liujunzhi@njnu.edu.cn (J. Liu).

https://doi.org/10.1016/j.ecolmodel.2018.03.005
Received 24 December 2017; Received in revised form 28 February 2018; Accepted 14 March 2018
Available online 30 March 2018
0304-3800/ © 2018 Elsevier B.V. All rights reserved.

considered homogeneous (Jones et al., 2003; Keating et al., 2003; Williams et al., 1989), i.e. represented by a single climate station. In order to assess the effect of climate change on crop growth over larger areas, some regional- or global-scale crop models emerged (Bondeau et al., 2007; Kucharik and Brye, 2003; Levis, 2010; Zhang et al., 2016). The Community Land Model (CLM) is one of the most widely-used land surface models, and has incorporated a comprehensive crop model, named CLM-Crop (Levis et al., 2012), which aims to simulate crop growth globally. This is helpful for predicting the impacts of climate change on agriculture over the world. However, the calibration and evaluation of CLM-Crop was only conducted in United State (Bilionis et al., 2015; Chen et al., 2015; Drewniak et al., 2013, 2015; Leng et al., 2016; Levis et al., 2012). Due to complex environment and human activities factors, it is not clear whether CLM-Crop, as a global-scale crop model, can be applied successful in other regions of the world.

In this study, we evaluate CLM-Crop for maize growth simulation in Northeast China, which is a commodity grain base with strategic role in China. The maize yield of this region accounts for 27% of China's total maize growing area, and over 30% of its total yield (National Bureau of Statistics of China (NBSC, 2011; Guo et al., 2013). Therefore, it is necessary to evaluate the suitability of CLM-Crop for maize growth modelling in this region in order to perform a reasonable global simulation. In this paper, Section 2 presents the descriptions of the study area, model setup and evaluation criteria. Section 3 evaluates the performance of CLM-Crop in simulating leaf area index (LAI), plant carbon, phenology and yield of maize using statistical data. Section 4 discusses the limitation of simulated phenology and yield, and proposes possible improvements to reduce simulation biases in this region. Finally, conclusions are presented in Section 5.

2. Methodology

2.1. Study area

Northeast China is located between 115.0°–135.0°E, 38.9°–53.0°N, including Heilongjiang, Jilin, Liaoning Provinces, and the east part of Inner Mongolia Autonomous Region. The Northeast Plain in this region is the largest plain of China, and consists of the Songnen, Liao river and Sanjiang Plains (Fig. 1). More than 90% of maize fields are rain-fed (National Bureau of Statistics of China (NBSC, 2011).

Climate is largely controlled by the East Asian monsoon which is characterized by warm and short summers, cold and dry winters. The mean annual air temperature is -4.2–10.9 ℃ and the annual accumulated temperature above 10 ℃ is 1700–3600 ℃, with the frost free period normally 155–257 days (Yao et al., 2015; Yin et al., 2016). The annual precipitation ranges from 400 to 1000 mm, and is mainly concentrated between July and September (Chen et al., 2012). The principal soils of Northeast China are Chernozems (mainly in the central section), Luvisols (mainly in the east), Cambisols (in the western mountain ranges) and Podzols (in the far north) (Institute of Soil Sciences, 1990), according to the World Reference Base for Soil Resources classification (FAO 2014).

2.2. Model setup

CLM-Crop is a part of the Community Earth System Model (CESM) (Dickinson et al., 2006; Lawrence et al., 2011). The current version (CLM 4.5) includes the simulation of spring wheat, maize and soybean which are represented as managed crop plant functional types (PFTs). The crop growth module in CLM-Crop is based on the AgroIBIS model (Kucharik and Brye, 2003), a well-known dynamic vegetation model.

Crop growth is divided into planting, emergence, grain filling, and harvest phases. The latter three phases are determined by the fraction of accumulated phenological heat units (FPHUs) (Drewniak et al., 2013) needed to achieve the respective phase. Planting date is determined by the 10-day running mean (T_{10d}) and minimum ($T_{10d\text{-}min}$) 2-m air temperature at every daily time step. During the emergence phase, carbon and nitrogen are allocated to leaves, stems, and roots, and LAI gradually increases and finally reaches its maximum which is prescribed constraints in CLM-Crop (a maximum LAI at 5.0 $m^2\,m^{-2}$). During the grain fill phase, the carbon accumulated in the leaves, stems and roots is reduced by translocation to the grain. Harvest occurs when the growing degree days (GDD) are equal to the growing degree-days necessary for the crop to reach physiological maturity (GDD_{mat}). Eq. (1) shows how GDD_{mat} is calculated by CLM-Crop:

$$GDD_{mat} = 0.85\ GDD_8 \quad (1)$$

Where GDD_8 is the 20-year running mean of accumulated GDD from April to September (in the northern hemisphere) base 8℃ (the maximum daily increments of 30°days). GDD_{mat} is bounded from 950 to 1850°days. So the growing period length can change over time. Detailed descriptions of CLM-Crop can be found in Oleson et al. (2013).

The input data of CLM-Crop are of two main types: meteorological forcing data and surface data. The meteorological forcing data, provided by the combined Climatic Research Unit and National Center for Environment Predication (CRUNCEP) dataset for 1901–2010, include

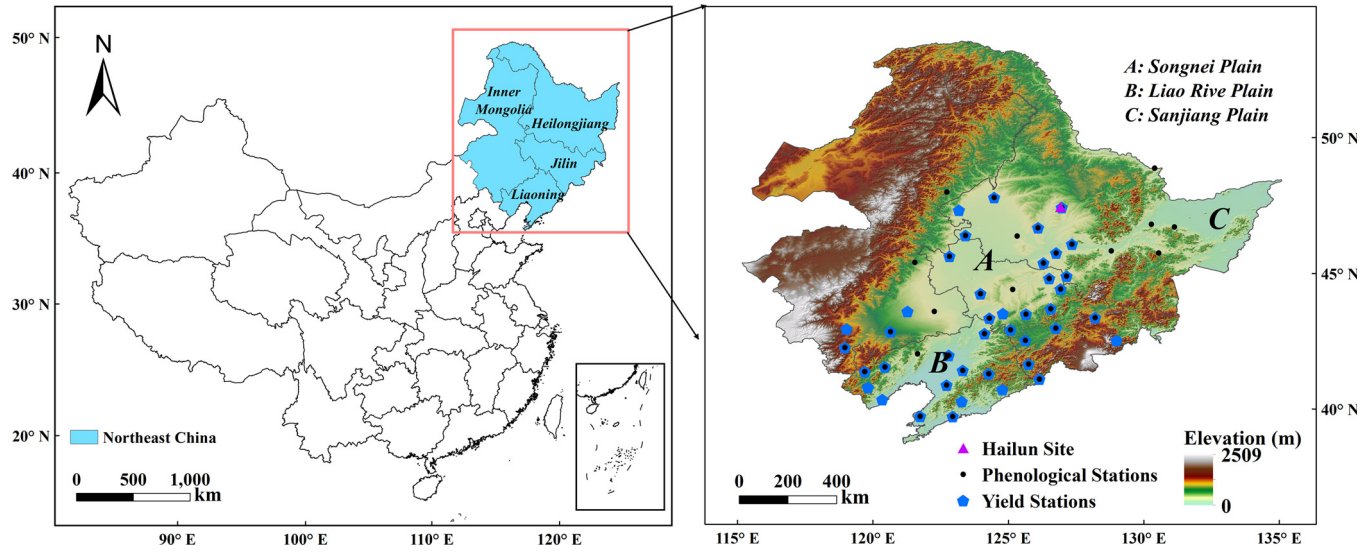

Fig. 1. Location of the study area and agricultural sites.

solar radiation, temperature, precipitation, wind speed, humidity, and surface pressure at 6-hour intervals (Viovy, 2011). The CRUNCEP dataset is a combination of two datasets which are the CRU TS3.2 0.5° × 0.5° grid resolution monthly data (1901–2002) (Mitchell and Jones, 2005) and the NCEP reanalysis 2.5° × 2.5° grid resolution 6-hourly data (1948–2010) (Kalnay et al., 1996). This dataset has been used to force CLM for numerous studies of evapotranspiration, vegetation growth, and gross primary production (Mao et al., 2013, 2012; Shi et al., 2013). Considering the computational load of the model, the resolution of the CRUNEP dataset, in this study CLM-Crop was run at a resolution of 0.5° × 0.5° (approximately 2180 km^2 at 45°N latitude), resulting in 606 pixels, for the period of 1991–2010.

In CLM-Crop, surface datasets assign the fractions of land types and PFTs in each grid cell. Crops are represented as unmanaged and managed crop PFTs. Each crop has its own soil column, and each soil column has its own carbon and nitrogen pools. Crops can coexist with natural vegetation in a grid cell, but they do not compete with natural vegetation for resources (Drewniak et al., 2013). The unmanaged crop dataset was derived from Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data (Lawrence and Chase, 2007) at 0.05°. The managed crop dataset in CLM-Crop was obtained from the present-day crop dataset of Portmann et al. (2010), which provides the global distribution and area of crops at 5 arc min. For consistency with the unmanaged crop dataset (MODIS-based, 0.05°, i.e., 3′), the managed crop dataset (5′) was resampled to 0.05° by bilinear interpolation. The managed crop PFTs were then assigned based on the managed crop dataset and unmanaged crop dataset for each grid (Leng et al., 2016). Finally, the land surface dataset was aggregated to the 0.5° grid resolution by summation. The distribution of fractions (%) of the maize growing area prescribed in the CLM-Crop model is shown in Fig. 2.

Before the real simulations, an accelerated spin-up procedure (Thornton and Rosenbloom, 2005) was used. During spin-up, CLM-Crop was run offline from 1991 to 2010 with CN cycling active. In order to ensure that the soil carbon and nitrogen pools reached a steady state, the forcing data were recycled 50 times (a total of 1000 years). After spin-up, the resulting state variables were used to initialize subsequent simulations. Then the CLM-Crop was run for 16 complete cycles of the 1991–2010 climate data (a total of 320 years). The last 20 years of simulation (one cycle of 1991–2010) was used for analysis with daily averaged data.

2.3. Evaluation of model performance

Three observational and statistical datasets were used to validate the simulation results. The first dataset was the observation data of LAI and plant carbon of maize in the State Key Agroecological Experiment Station, Hailun County, Heilongjiang Province, China (47°26′N, 126°38′E) (Fig. 1). The second dataset was the phenology and yield data at agricultural meteorological stations of the Chinese Meteorological Administration (CMA archives, 2010), including 43 phenological stations and 41 yield stations (Fig. 1). These stations have uniform observing guidelines and standards of CMA for collecting phenology and crop yield data (CMA archives, 2010; Liu et al., 2012). There are 691, 688 and 606 observed samples for planting date (PD), harvesting date (HD) and growing period (GP) from 1993 to 2010 and 377 observations for maize yield from 1999 to 2010, respectively. The third dataset was the annual county yield data from the National Bureau of Statistics of China (NBSC), which was collected by local investigators from field sampling using uniform agricultural production sampling survey system across the county (National Bureau of Statistics of China (NBSC, 2011). In this study, yield data for 174 counties and 20 years (1991–2010) were used to validate. The areas of these counties range from 740 to 55 096 km^2, with a median of 4268 km^2. The total study is 131 × 10^4 km^2, corresponding to 606 grid cells (0.5° resolution), so that most counties were represented by one to four grid cells.

The correlation coefficient (r) and the root mean square error (RMSE) were used to evaluate model performance (Eqs. (2) and (3)) (Schaefer et al., 2012).

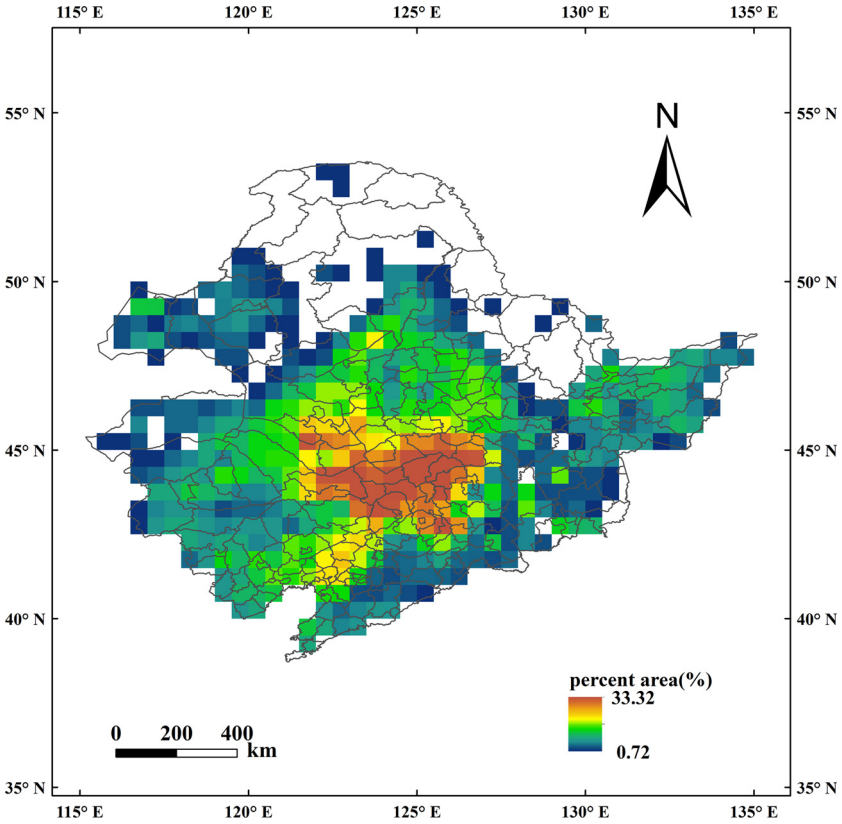

Fig. 2. The spatial distribution of the maize growing area percentage relative to the grid area.

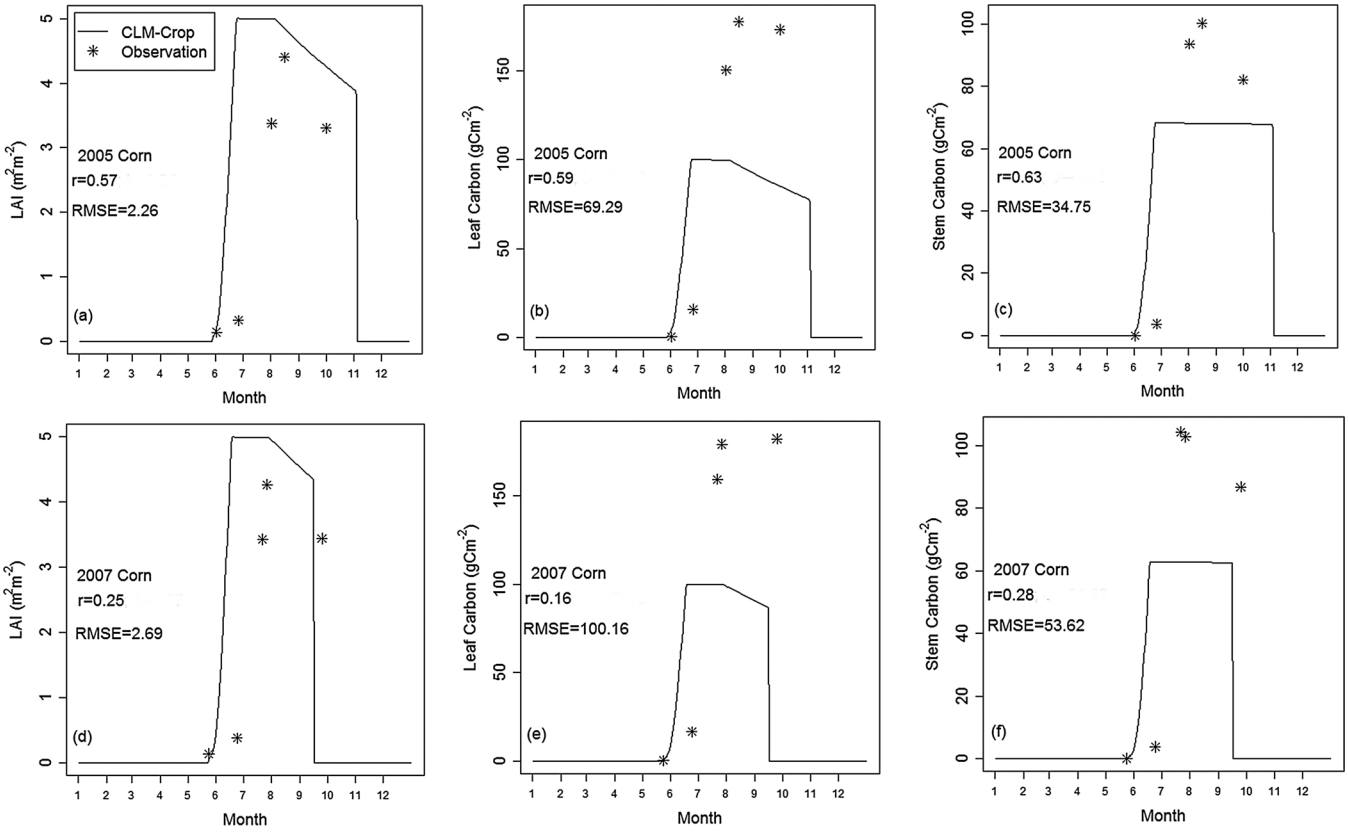

Fig. 3. Simulated LAI, leaf and stem carbon (shown with solid line) compared with field observations (asterisks).

$$r = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \quad (3)$$

where y_i is the observed or statistical values, $\hat{y}_i$ is the simulated values, the overbars represent the means, and n is the number of samples. If r is close to 1, it suggests that the model performs well. The RMSE is calculated as the standard deviation of the residuals against a 1:1 relation, so that RMSE close to 0 means good performance.

3. Results

3.1. LAI and plant carbon

Observation data of the Hailun site in 2005 and 2007 was used to validate the simulated LAI and plant carbon against modelled values for the grid cell where the Hailun site is located (center of pixel 47.25 °N, 126.75 °E).

Simulated (lines) and observed (asterisks) LAI are shown in Fig. 3 (a, d). CLM-Crop generally overestimated LAI during the growing period in these two years. The average correlation coefficient was 0.41, and the average RMSE was 2.47. Previous studies have also observed overestimated LAI in CLM3.5-Crop (Drewniak et al., 2013) and CLM4.5-Crop because simulated leaf emergence was earlier (Chen et al., 2015; Peng et al., 2018). Fig. 3(a, d) shows that the simulated LAI reached a maximum more quickly than observation and before leaf senescence the simulated LAI maintained constant maximum LAI (Peng et al., 2018). In addition, the LAI in the model is calculated using a fixed value for specific leaf area (SLA; the ratio of leaf area to dry leaf weight), but SLA actually varies in the growth period (Drewniak et al., 2013; Tardieu et al., 1999). This is also the possible reason of overestimated LAI.

Fig. 3 (b, c, e, f) shows the simulated and observed plant carbon. Overall, both the leaf and stem carbon were underestimated by the CLM-Crop model. The average correlation coefficients and RMSEs were 0.38 and 84.72 g C m^{-2} for leaf carbon, and 0.46 and 44.18 g C m^{-2} for stem. The maximum simulated leaf and stem carbon were only 100.10 g C m^{-2} and 68.25 g C m^{-2}, respectively, which were much lower than the maximum observed value (182.50 and 104.40 g C m^{-2}, respectively). The simulated leaf and stem carbon increased fast during the early growing period and reached maximum values one month earlier than the observations. It is worth noting that in the CLM-Crop, when the prescribed maximum LAI is reached, all available carbon is allocated from above ground components (e.g. leaf, stem) to roots. But there is little theoretical basis for this approach (Peng et al., 2018).

3.2. Phenology

The simulated planting and harvesting dates, as well as the length of growing period, were compared with observation data from 43 agricultural meteorological stations from 1993 to 2010. In general, the median values for simulations and observations were close. The simulated dates were all slightly later than observation values for planting and harvesting date, and longer for growing period (Fig. 4). The simulated values had larger ranges than observation values, especially for harvesting dates. The simulation results for planting dates were better than those for harvesting dates. The correlation coefficients between simulation and observation was 0.53 (RMSE = 9.22 days) for planting dates, and 0.20 (RMSE = 18.72 days) for harvesting dates. For growing periods, the correlation coefficient was only 0.04 (RMSE = 17.34 days).

Fig. 5 shows the differences between mean simulated and observed planting/harvesting dates and growing period in the 43 agricultural meteorological stations. About 70% of the stations had differences within one week between the simulated and observed planting dates

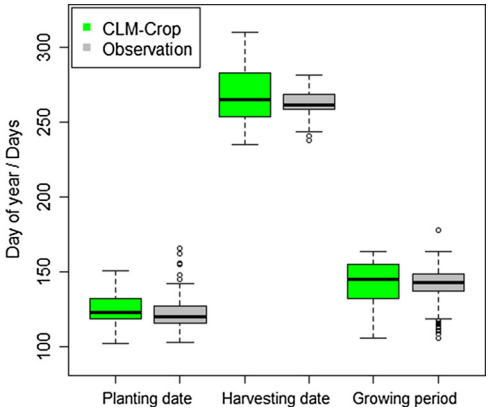

Fig. 4. Simulation (green) and observation (grey, 43 stations) values for maize planting date (day of year), harvesting date (day of year) and growing period (days) in Northeast China from 1993 to 2010 (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

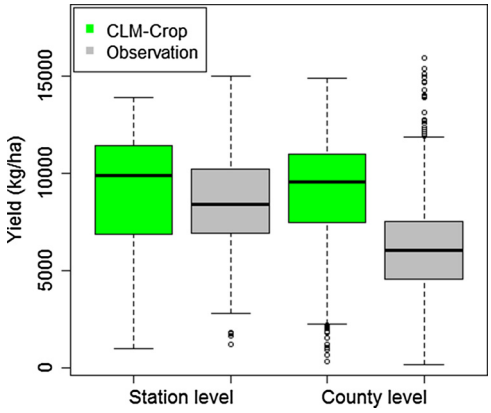

Fig. 6. CLM-Crop-simulated (green) and observed (grey) maize yields (kg/ha) in Northeast China at the station and county level (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

(Fig. 5a). For harvesting dates (Fig. 5b), about 60% of the stations had differences within one week. Those sites with small differences were mainly distributed in the middle plain region of Northeast China for both planting and harvesting dates (Fig. 5a, b). For planting dates, the simulation dates were later than the actual dates in the east part of the region, while the opposite is the case in the southwest part. For harvesting dates, the simulation dates were later than observation values in most of the sites except for those in the middle plain region. Several harvest dates were severely overestimated (up to one month). For growing periods, about 42% of the stations were underestimated which were mainly distributed in the middle plain region. The overestimated stations were mainly distributed in the surrounding mountainous areas (Fig. 5c). Part of the discrepancy is because in the CLM-Crop, the planting date and harvesting date are determined only by temperature (Section 2.2), but in practice they are determined by farmers' choice and operational limitations (Sacks et al., 2010).

3.3. Yields

Fig. 6 shows the comparison between simulated and observed maize yields over Northeast China. Observation data from 41 agricultural meteorological stations from 1999 to 2010 (a total of 377 observations) were used for validation. Simulated yields ranged from 993 to 13,905 kg/ha (mean 9066), while the observed yields ranged from 1184 to 14,978 kg/ha (mean 8561), thus a fairly close match, with simulated mean yield slightly larger than observed mean yield. At county level, the statistical yield from 174 counties (1991–2010, 2620 samples) were used for validation. The simulated yields ranged from 317 to 14,880 kg/ha (mean 9091), and the observed yields ranged from 150 to 15,920 kg/ha (mean 6098). The mean simulated yield was much larger than the mean county statistical yield.

Fig. 7a shows a scatterplot of CLM-Crop-simulated and observed maize yields at the station level, which indicates a poor correlation between simulation and observation ($r = -0.15$, RMSE = 4036 kg/ha). However, the simulation and observation values had much better correlation ($r = 0.26$, RMSE = 4213 kg/ha) at the county level, although there was an overall overestimation (Fig. 7b). Some reported county yields exceed the corresponding model simulations; this may be due to over-optimistic or unrepresentative reporting. The overall overestimation is typical of simulated vs. farmer yields. It is well-known that a major cause of mismatch between simulated and actual yields is because of management decisions by individual farmers, including suboptimal timing of farming operations, inappropriate variety selection, and poor weed and pest control. Many studies compensate for this by a so-called yield bias correction. As a typical example, Jagtap and Jones (2002) computed these corrections for a 17-year period in their study of regional yields of soybeans simulated by the CROPGRO model, and found that these were in the range of 0.4–0.5 in about three-quarters of the 58 (0.5° resolution) grid cells covering Georgia (USA). So we were

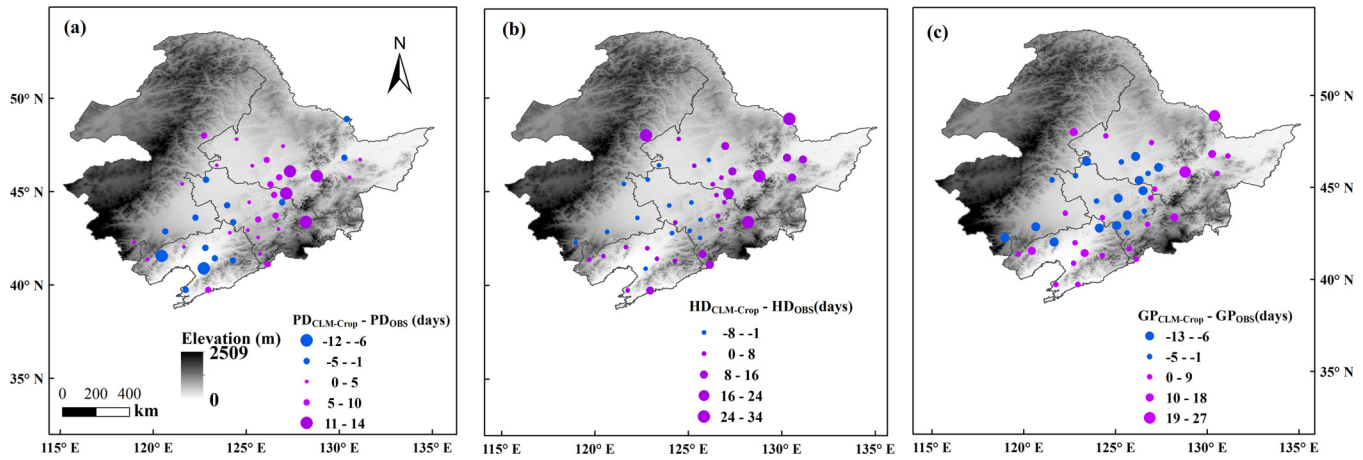

Fig. 5. The differences between mean simulated and observed planting dates (PD, a), harvesting dates (HD, b) and growing period (GP, c) in the 43 agricultural meteorological stations over 1993–2010. The location of the stations is shown as points, whose color and size represent the difference in days between simulated and observed values. The digital elevation model is shown as background.

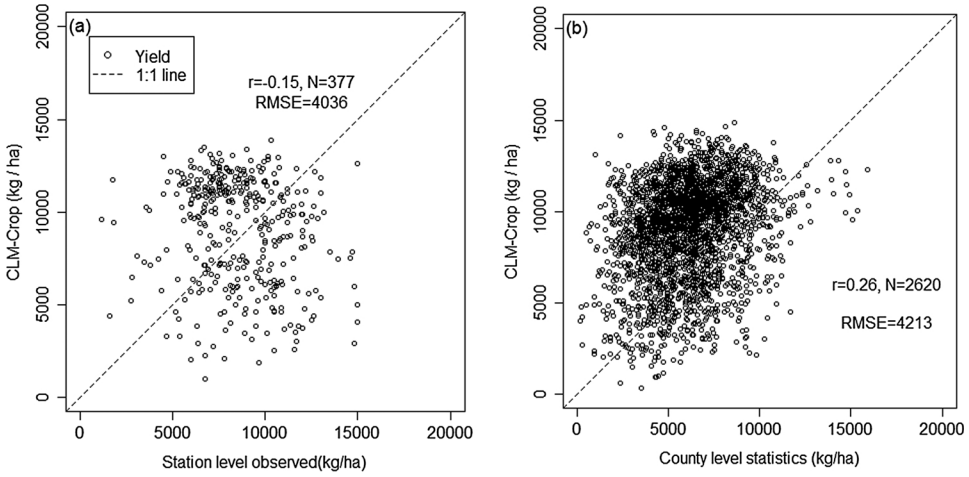

Fig. 7. Scatterplot of CLM-Crop-simulated and observed/statistical maize yields at the station-level (a) and county-level (b).

not surprised that the median county-average yield of 6055 kg/ha is about 40% below the median simulated county yield of 9573 kg/ha.

Fig. 8 shows the spatial distribution of simulated (a) and statistical (b) average maize yield from 1991 to 2010, and their differences (c) at the county level. The simulated result (Fig. 8a) shows maize yield decreasing from the southeast to the northwest, corresponding mainly to shorter growing seasons. However, the statistical average yield shows a different spatial pattern (Fig. 8b). The areas with high yields were mainly concentrated in the south part of the Songnei plain (labelled "A" in Fig. 1), followed by other plain areas, and low yields were mainly distributed in surrounding mountainous areas. In most counties, the simulated yields were larger than statistical yields (Fig. 8c), with the highest differences occurring in the south-east mountainous areas.

4. Discussion

4.1. Uncertainties and limitations of simulated phenology

Phenology simulation plays an important role in modeling crop growth, and planting date is a basic variable in phenology simulation (Bilionis et al., 2015; Deryng et al., 2011; Sacks et al., 2010; Waha et al., 2012). Drewniak et al. (2013) indicated that planting date is one of the most sensitive variables in CLM-Crop, which has substantial influence on growth period and the resulting carbon fluxes. In the middle plains region of Northeast China, the simulated planting dates agree well with observations. But in surrounding mountainous areas, the simulated planting dates were about two weeks later than observations (Fig. 5). This may be because in complex terrains in northern latitudes, most cropland may be located in valleys or on south-facing slopes to take advantage of warmer conditions. As the maize growing areas are concentrated in plains other than in mountains, the low performance in mountainous areas may not have much negative impacts. But in order to obtain more reasonable spatial distribution of phenology and yield, it is still necessary to improve the parameterization of CLM-Crop in regions with complex terrain, either by increasing the resolution of the climate data set or adjusting it according to cropland distribution within each grid.

In CLM-Crop, the planting date is calculated according to crop-specific temperature thresholds, i.e., the coldest temperatures at which planting can occur. When T_{10d} is greater than 10 °C and $T_{10d\text{-}min}$ is greater than 6 °C, maize can be planted. However, the actual planting dates recorded for the 43 stations show that the temperature thresholds of planting are not fixed but vary spatially. The spatial distribution of actual T_{10d} and $T_{10d\text{-}min}$ which were calculated based on observation planting dates from 43 stations (1993–2010) are shown in Fig. 9. In most of the phenology stations, the actual T_{10d} was less than 10 °C (Fig. 9a) and the actual $T_{10d\text{-}min}$ was more than 6 °C (Fig. 9b). It suggests that the temperature thresholds of planting are variable in different

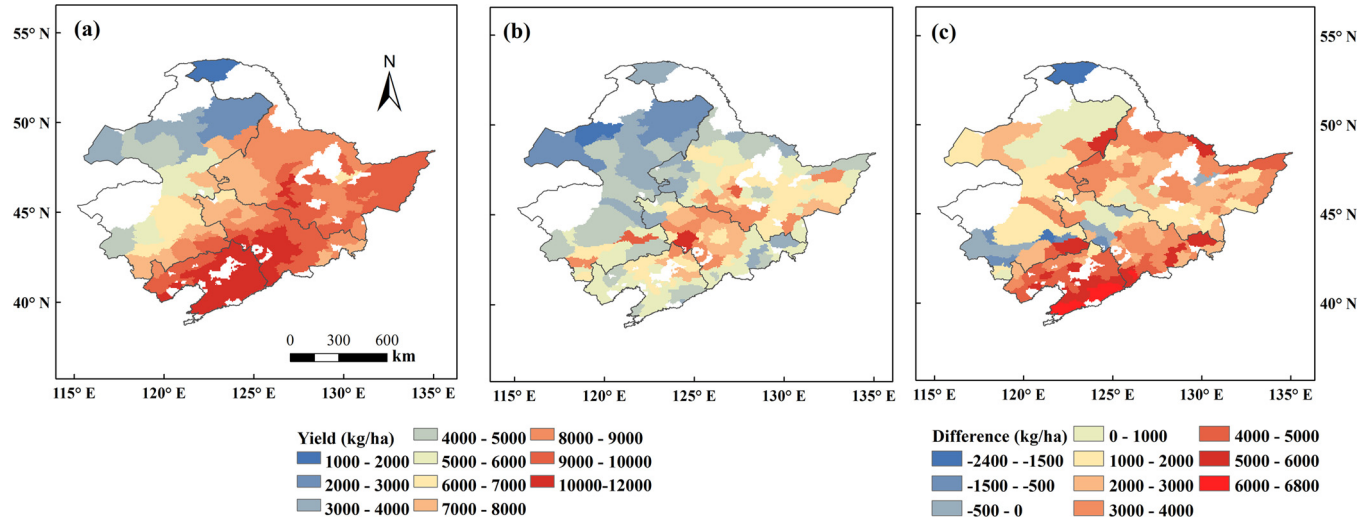

Fig. 8. The spatial distribution of simulated (a) and statistical (b) average maize yield from 1991 to 2010 at county-level. The spatial distribution of the differences between simulated and statistical maize yield (c).

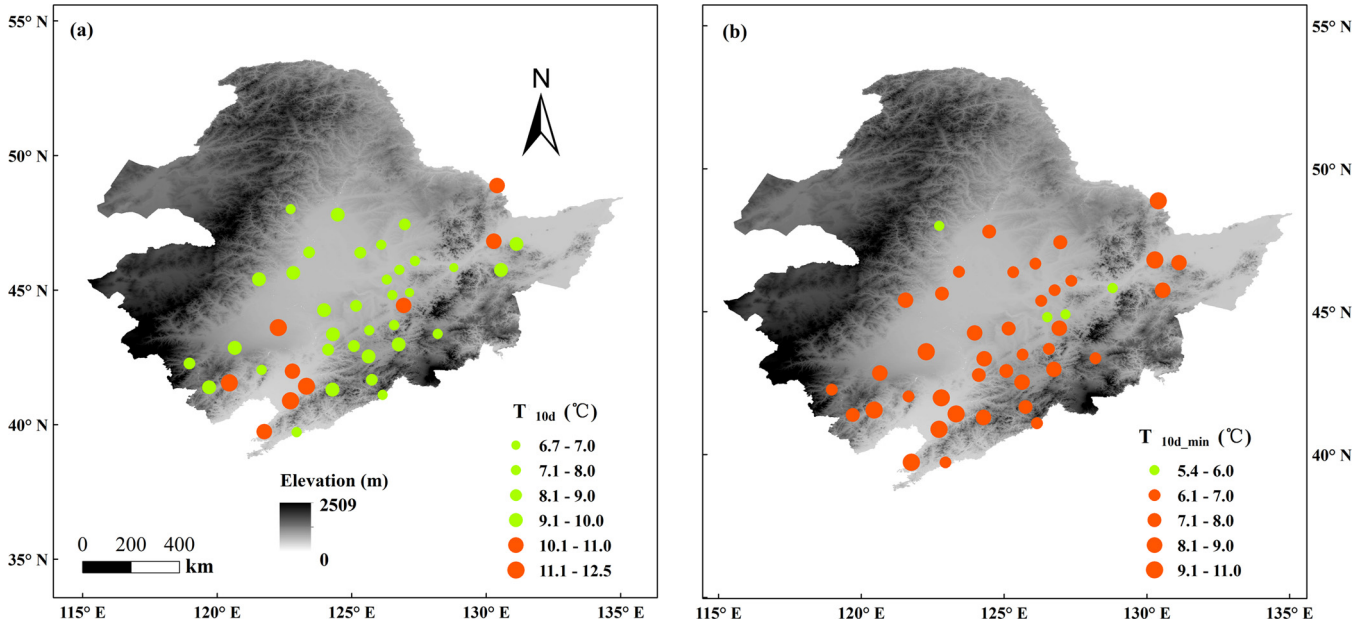

Fig. 9. The spatial distribution of the actual T_{10d} (a) and $T_{10d\text{-}min}$ (b) calculated based on observed planting dates in 43 agricultural meteorological stations (1993-2010).

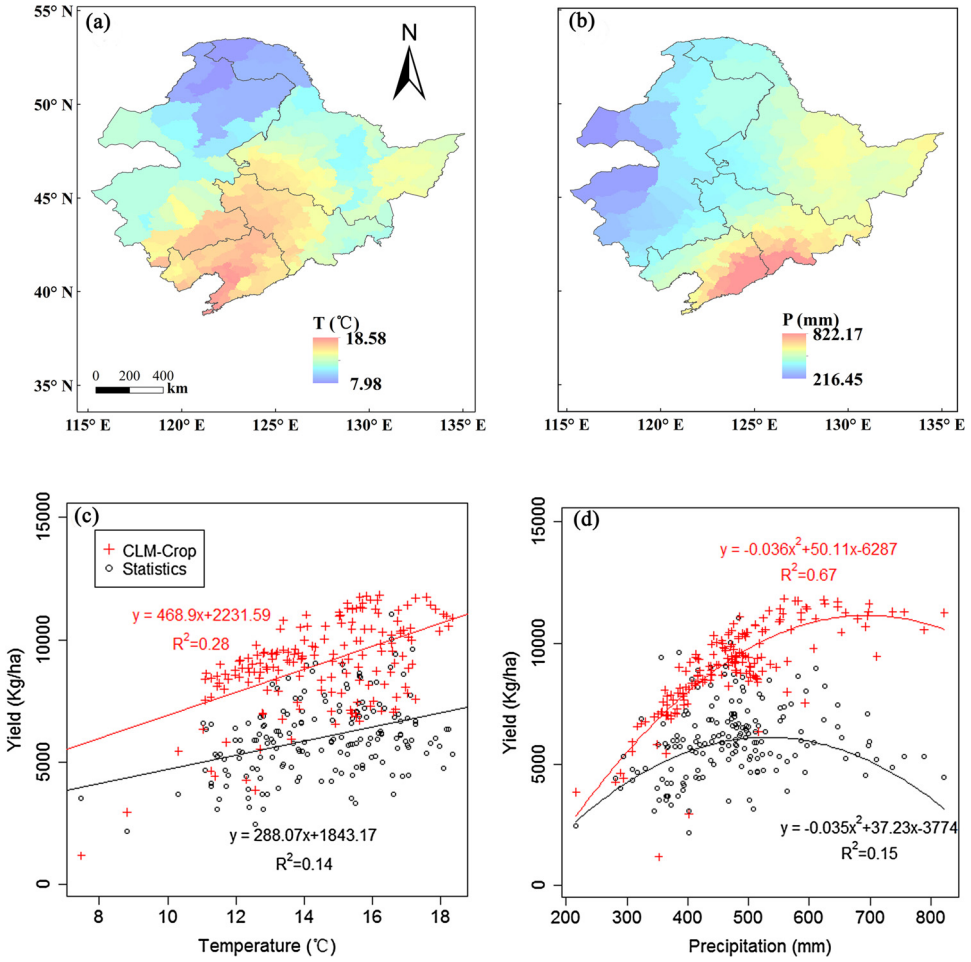

Fig. 10. The spatial distribution of mean temperature (T, a), precipitation (P, b) averaged over 1991–2010 during the maize growing season (April through October). The bottom panels are the relationships between simulated/statistical yields and mean temperature (c), precipitation (d).

place, perhaps because of different varieties or microclimate, and it is unreasonable that the fixed temperature thresholds were used in CLM-Crop at regional or global scale.

Therefore, this is likely to constitute one source of error in modeling phenology. Planting date is usually determined by farmers' choice, not only considering temperature, but also precipitation, soil moisture, cultivar variety, agronomic technologies and other conditions favorable for growth (Kucharik, 2006; Sacks et al., 2010). So the threshold of planting is triggered not only by the temperature but also by the terrain, human activities such as farming practices and other factors. It is necessary to improve the understanding of crop phenology and develop gridded crop parameters to improve the model performance at the regional or global scale.

4.2. Uncertainties and limitations of simulated yield of maize

In this study, station-level observed yields and the county-level statistical yields were used to validate the simulation results of CLM-Crop. Although the average simulated maize yield was close in absolute terms (RMSE) to station-level observations, the correlation between simulated and observed yield was very poor (Fig. 7a). At the county level, the simulated and statistical yields had much better correlation although there was an overall overestimation (Fig. 7b). The poor correlation between simulated and observed maize yield at the station level may be related to scale mismatch. In this study, CLM-Crop was run at a resolution of $0.5^{\circ} \times 0.5^{\circ}$, which cannot be represented by a single station except in very uniform agricultural landscapes. At the county level, the scales of simulated and statistical yields were more consistent because a county covers one or several grid cells. So the county statistical yields can to some extent represent a regional ground truth and are more suitable for the validation of simulated crop yields (Chen et al., 2011; Mo et al., 2005; Yao et al., 2015).

Crop yields are seriously affected by the temperature, precipitation and agricultural management practices such as tillage, weed control and irrigation. To investigate the extent to which these factors may have affected the yield discrepancies, we used the mean temperature and precipitation as factors to analyze the relationship between these factors and simulated or statistical yield. The mean temperature and precipitation during the maize growing season (April through October) were averaged over 1991–2010. The mean temperature and precipitation are independent variables.

Fig. 10(a, b) shows the spatial distribution of mean temperature and precipitation. Fig. 10(c, d) show scatterplots of statistical and simulated yields for each county, with best-fit polynomial regression equations selected by adjusted R^2 and Akaike Information Criterion (AIC). The spatial distribution of simulated yields was consistent with that of temperature and precipitation (Fig. 8a, Fig. 10a, b), with the obvious positive bias of simulated yields discussed above, and a slightly stronger response to temperature (steeper slope, doubled R^2) for simulated yields. This suggests that CLM-Crop can well capture the effect of the temperature.

However, the relation with precipitation is not monotonic (Fig. 10d), especially for the statistical yields. For simulated yields, when the precipitation increased to about 750 mm, the simulated yield reaches the maximum value and slightly decreases; this is interpreted as a threshold effect, where precipitation is no longer the limiting factor. But for statistical yields, when the precipitation increased to 500 mm, the statistical yield reaches the maximum value and then decreases, although because of the scatter and fairly few observations with such high precipitation, this may also be a threshold effect, but at a lower yield threshold. This may be because limitations in farmer management is not able to take full advantage of the available precipitation. Also striking is that the fitted polynomial is much steeper at low to medium precipitations for the simulated, as opposed to statistical yields. This also suggests that farmer practice is not taking full advantage of available precipitation, perhaps because of weed competition or improper planting densities.

Drewniak et al. (2013) noted that the simulated yields from CLM-Crop have a strong positive correlation with precipitation, consistent with our results for simulated yields. Accordingly, in the south of Northeast China with the precipitation of about 800 mm, the simulated maize yield is more than other areas (Fig. 10b, Fig. 8a), which is a serious overestimation. On the contrary, in the middle plain areas with moderate precipitation, the simulated yields were comparable with statistical yields, taking into account the overall bias explained above. In sum, CLM-Crop can capture the effect of the temperature and precipitation on crop yields. The impact of precipitation on actual yields was exaggerated in CLM-Crop. It is necessary to adjust CLM-Crop to properly reflect the influence of precipitation, taking into account other factors that reduce the effect of precipitation on yield. As Fig. 10d shows, this is not simply a bias adjustment.

As just mentioned, management could be the reason for much of the disagreement between simulated and statistical yields. This include uncertainties in the CLM-Crop management practices. For example, in CLM-Crop, fixed fertilizer rates and higher fertilizer than typically used in cultivation will lead to overestimated yields (Leng et al., 2016). Therefore, for realistic application, gridded crop management parameters should be developed based on field or statistical surveys to determine which cultivars are grown, what fertilizers are used, etc., at the resolution of the CLM-Crop grid. One example is that Potter et al. (2010), who established the spatial distribution of global fertilizer application by crop type using fertilizer datasets.

5. Conclusion

As a state-of-the-art process-based land surface model, CLM-Crop is designed to be applied at a single site, region scale or global scale. In this study, we evaluated the CLM-Crop through validating for LAIs, plant carbons, phenology and yields over the Northeast China and proposed possible improvements. Despite overall good results, CLM-Crop overestimated LAIs, and underestimated leaf and stem carbon for maize. For simulation of phenology and yield, the CLM-Crop performed well in the middle plain areas. In the mountains areas, the performance of CLM-Crop should be improved by increasing the resolution of the climate and surface dataset to better characterize the climate and terrain conditions of crop growing areas. The county level statistical yields better corresponded to the simulated spatial pattern of yields than station level, perhaps because CLM-Crop was run with 0.5° resolution climate information and not at near-point resolution corresponding to experiment stations. CLM-Crop was able to capture the effect of the temperature on crop yields. But CLM-Crop excessively considered the effect of precipitation on crop yields, especially in moderately high precipitation areas. If CLM-Crop is applied to scenarios such as extreme climate events, especially heavy precipitation, this result could be amplified.

Thus, the following improvements could be made to CLM-Crop, to give more realistic results in Northeast China. 1) In complex terrains, it is necessary to consider the impacts of microclimate on crop phenology and yield. 2) CLM-Crop should be adjusted to properly reflect the influence of precipitation on yield. 3) Gridded crop-specific phenological and management parameters should be developed in order to improve the model performance at regional scale.

Acknowledgments

This work was supported by National Basic Research Program of China (Project No.: 2015CB954102), Natural Science Foundation of Jiangsu Province of China (Project No.: BK20150975), National Natural Science Foundation of China (Project No.: 41601413, 41431177), Natural Science Research Program of Jiangsu (Project No.: 14KJA170001), and PAPD (Project No.: 164320H116). Supports to A-Xing Zhu through the Vilas Associate Award, the Hammel Faculty

Fellow Award, the Manasse Chair Professorship from the University of Wisconsin-Madison, and the “One- Thousand Talents” Program of China are greatly appreciated. Supports from D.G. Rossiter through the NJNU Visiting Prof. Programme. Supported by Program of Innovative Research Team of Jiangsu Higher Education Institutions of China.

References

Bilionis, I., et al., 2015. Crop physiology calibration in the CLM. Geosci. Model Dev. 8 (4), 1071–1083.

Bondeau, A., et al., 2007. Modelling the role of agriculture for the 20th century global terrestrial carbon balance. Global Change Biol. 13 (3), 679–706.

Chen, C., et al., 2011. Will higher minimum temperatures increase corn production in Northeast China? An analysis of historical data over 1965–2008. Agric. For. Meteorol. 151 (12), 1580–1588.

Chen, C., et al., 2012. Progressive and active adaptations of cropping system to climate change in Northeast China. Eur. J. Agron. 38, 94–103.

Chen, M., et al., 2015. Simulating crop phenology in the community land model and its impact on energy and carbon fluxes. J. Geophys. Res. Biogeosci. 120 (2), 310–325.

CMA archives, 2010. Chinese Meteorological Administration Archives. http://data.cma.cn/.

Deryng, D., et al., 2011. Simulating the effects of climate and agricultural management practices on global crop yield. Global Biogeochem. Cycles 25 (2), 96–101.

Deryng, D., et al., 2014. Global crop yield response to extreme heat stress under multiple climate change futures. Environ. Res. Lett. 9 (3), 034011.

Dickinson, R.E., et al., 2006. The community land model and its climate statistics as a component of the community climate system model. J. Clim. 19 (11), 2302–2324.

Drewniak, B., et al., 2013. Modeling agriculture in the community land model. Geosci. Model Dev. 6 (4), 495–515.

Drewniak, B.A., et al., 2015. Modeling the impact of agricultural land use and management on US carbon budgets. Biogeosci. 12 (7), 2119–2129.

Ed, H., et al., 2013. Increasing influence of heat stress on French maize yields from the 1960s to the 2030s. Global Change Biol. 19 (3), 937–947.

Estes, L.D., et al., 2013. Projected climate impacts to South African maize and wheat production in 2055: a comparison of empirical and mechanistic modeling approaches. Global Change Biol. 19 (12), 3762–3774.

Fader, M., et al., 2010. Virtual water content of temperate cereals and maize: present and potential future patterns. J. Hydrol. 384 (3–4), 218–231.

FAO, 2014. World Reference Base for Soil Resources 2014: International Soil Classification System for Naming Soils and Creating Legends for Soil Maps. Retrieved from. Food and Agriculture Organization of the United Nations, Rome. http://www.fao.org/3/a-i3794e.pdf.

Guo, J., et al., 2013. Evaluation of agricultural climatic resource utilization during spring maize cultivation in Northeast China under climate change. J. Meteorolog. Res. 27 (5), 758–768.

He, L., et al., 2015. Impacts of recent climate warming, cultivar changes, and crop management on winter wheat phenology across the Loess plateau of China. Agric. For. Meteorol. 200, 135–143.

Hoffman, A.L., et al., 2018. Analysis of climate signals in the crop yield record of sub-Saharan Africa. Global Change Biol. 24 (1), 143–157.

Institute of Soil Science, Chinese Academy of Sciences, 1990. In: Hong, Q., Chen, P., Li, C. (Eds.), Soils of China. Science Press, Beijing.

Intergovernmental Panel on Climate Change (IPCC), 2007. Climate change: impacts, adaptations and vulnerability. In: Parry, M.L., Canziani, O.F., Palutikof, J.P., can der Linden, P.J., Hanson, C.E. (Eds.), Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change.

Jagtap, S.S., Jones, J.W., 2002. Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. Agric. Ecosyst. Environ. 93 (1), 73–85. http://dx.doi.org/10.1016/S0167-8809(01)00358-9.

Jones, J.W., et al., 2003. The DSSAT cropping system model. Eur. J. Agron. 18 (3), 235–265.

Kalnay, E., et al., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77 (3), 437–472.

Keating, B.A., et al., 2003. An overview of APSIM, a model designed for farming systems simulation. Eur. J. Agron. 18 (3), 267–288.

Kucharik, C.J., 2006. A multidecadal trend of earlier corn planting in the central USA. Agron. J. 98, 1544–1550.

Kucharik, C.J., Brye, K.R., 2003. Integrated BIosphere simulator (IBIS) yield and nitrate loss predictions for Wisconsin maize receiving varied amounts of nitrogen fertilizer. J. Environ. Qual. 32 (1), 247.

Lawrence, D.M., et al., 2011. Parameterization improvements and functional and structural advances in version 4 of the community land model. J. Adv. Model. Earth Syst. 3 (3).

Lawrence, P.J., Chase, T.N., 2007. Representing a MODIS consistent land surface in the community land model (CLM3.0). J. Geophys. Res. 112, G01023. http://dx.doi.org/10.1029/2006JG000168.

Leng, G., et al., 2016. Simulating county-level crop yields in the conterminous united states using the community land model: the effects of optimizing irrigation and fertilization. J. Adv. Model. Earth Syst. 8 (4), 1912–1931.

Levis, S., et al., 2012. Interactive crop management in the Community Earth System Model (CESM1): Seasonal influences on land–atmosphere fluxes. J. Clim. 25 (14), 4839–4859.

Levis, S., 2010. Modeling vegetation and land use in models of the earth system. Wiley Interdiscip. Rev. Clim. Change 1 (6), 840–856.

Liu, J., et al., 2007. GEPIC–modelling wheat yield and crop water productivity with high resolution on a global scale. Agric. Syst. 94 (2), 478–493.

Liu, Z., et al., 2012. Maize potential yields and yield gaps in the changing climate of northeast China. Global Change Biol. 18 (11), 3441–3454.

Lobell, D.B., Asseng, S., 2017. Comparing estimates of climate change impacts from process-based and statistical crop models. Environ. Res. Lett. 12 (1), 015001.

Lobell, D.B., Ortizmonasterio, J.I., 2007. Impacts of day versus night temperatures on spring wheat yields. Agron. J. 99 (2), 469–477.

Lobell, D.B., et al., 2012. Extreme heat effects on wheat senescence in India. Nat. Clim. Change 2 (3), 186–189.

Lobell, D.B., et al., 2014. Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. Science 344 (6183), 516–519.

Mao, J., et al., 2012. Remote sensing evaluation of CLM4 GPP for the period 2000–09. J. Clim. 25 (15), 5327–5342.

Mao, J., et al., 2013. Global latitudinal-asymmetric vegetation growth trends and their driving mechanisms: 1982–2009. Remote Sens. 5 (3), 1484–1497.

Mitchell, T.D., Jones, P.D., 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int. J. Climatol. 25 (6), 693–712.

Mo, X., et al., 2005. Prediction of crop yield, water consumption and water use efficiency with a SVAT-crop growth model using remotely sensed data on the North China plain. Ecol. Modell. 183 (2–3), 301–322.

National Bureau of Statistics of China (NBSC), 2011. China Statistics Yearbook. China Statistics Press, Beijing.

Oleson, K.W., Lawrence, D.M., Bonan, G.B., et al., 2013. Technical description of version 4.5 of the Community Land Model (CLM). Ncar Technical Note NCAR/TN-503 + STR. National Center for Atmospheric Research, Boulder pp. 422.

Osborne, T., et al., 2009. Examining the interaction of growing crops with local climate using a coupled crop–climate model. J. Clim. 22 (6), 1393–1411.

Peng, B., et al., 2018. Improving maize growth processes in the community land model: implementation and evaluation. Agric. For. Meteorol. s 250–251, 64–89.

Peng, S., et al., 2004. Rice yields decline with higher night temperature from global warming. Proc. Natl. Acad. Sci. U.S.A 101 (27), 9971–9975.

Piao, S., et al., 2010. The impacts of climate change on water resources and agriculture in China. Nature 467 (7311), 43–51.

Portmann, F.T., et al., 2010. MIRCA2000-global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling. Global Biogeochem. Cycles 24 (1), GB1011.

Potter, P., et al., 2010. Characterizing the spatial patterns of global fertilizer ap-plication and manure production. Earth Interact. 14, 1–22. http://dx.doi.org/10.1175/2009EI288.1.

Rosenzweig, C., et al., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. PNAS 111 (9), 3268–3273.

Roudier, P., et al., 2011. The impact of future climate change on West African crop yields: what does the recent literature say? Global Environ. Change 21 (3), 1073–1083.

Sacks, W.J., et al., 2010. Crop planting dates: an analysis of global patterns. Global Ecol. Biogeogr. 19 (5), 607–620.

Schaefer, K., et al., 2012. A model-data comparison of gross primary productivity: results from the North American carbon program site synthesis. J. Geophys. Res. Biogeosci. 117 (G3).

Schlenker, W., Roberts, M.J., 2009. Nonlinear temperature effects indicate severe damages to US crop yields under climate change. PNAS 106 (37), 15594–15598.

Shi, X., et al., 2013. Spatiotemporal patterns of evapotranspiration in response to multiple environmental factors simulated by the community land model. Environ. Res. Lett. 8 (2), 024012.

Siebert, S., et al., 2014. Impact of heat stress on crop yield—on the importance of considering canopy temperature. Environ. Res. Lett. 9 (4), 044012.

Tardieu, F., et al., 1999. Modelling leaf expansion in a fluctuating environment: are changes in specific leaf area a consequence of changes in expansion rate? New Phytol. 143 (1), 33–43.

Thornton, P.E., Rosenbloom, N.A., 2005. Ecosystem model spin-up: estimating steady state conditions in a coupled terrestrial carbon and nitrogen cycle model. Ecol. Modell. 189 (1–2), 25–48.

Viovy, N. 2011. CRUNCEP dataset. [Description available at http://dods.extra.cea.fr/data/p529viov/cruncep/readme.htm. Data available at http://dods.extra.cea.fr/store/p529viov/cruncep/V4_1901_2011/].

Waha, K., et al., 2012. Climate-driven simulation of global crop sowing dates. Global Ecol. Biogeogr. 21 (2), 247–259.

Wang, P., et al., 2011. Yield estimation of winter wheat in the North China plain using the remote-sensing–photosynthesis–yield estimation for crops (RS–P–YEC) model. Int. J. Remote Sens. 32 (21), 6335–6348.

Williams, J., et al., 1989. The EPIC crop growth model. Trans. ASAE 32 (2) 497–0511.

Yao, F., et al., 2015. Estimation of maize yield by using a process-based model and remote sensing data in the Northeast China plain. Phys. Chem. Earth Parts A/B/C 87, 142–152.

Yin, X., et al., 2016. Impacts and adaptation of the cropping systems to climate change in the Northeast farming region of China. Eur. J. Agron. 78, 60–72.

Zhang, L., et al., 2016. Evaluation of the community land model simulated carbon and water fluxes against observations over ChinaFLUX sites. Agric. For. Meteorol. 226–227, 174–185.

Zhao, G., et al., 2013. Large-scale, high-resolution agricultural systems modeling using a hybrid approach combining grid computing and parallel processing. Environ. Modell. Softw. 41, 231–238.